A hybrid particle-stochastic map filter

Pengcheng Hao [a],[*], Oktay Karakuş [b], Alin Achim [a]

[a] *Visual Information Laboratory, University of Bristol, Bristol, UK*
[b] *School of Computer Science and Informatics, Cardiff University, Cardiff, UK*

ARTICLE INFO

ABSTRACT

Filtering in nonlinear state-space models is known to be a challenging task due to the posterior distribution being either intractable or expressed in a complex form. One of the most successful methods, particle filtering (PF), although generally outperforming traditional filters, suffers from sample degeneracy. Drawing from optimal transport theory, the stochastic map filter (SMF) accommodates a solution to this problem, but its performance is influenced by the limited flexibility of nonlinear map parameterisation. To alleviate these drawbacks, we propose a hybrid filter which combines the PF and SMF, and hence call it PSMF. Specifically, the PSMF splits the likelihood into two parts, which are then updated by PF and SMF, respectively. The proposed approach adopts systematic resampling and smoothing to break the particle degeneracy caused by the PF. To investigate the influence of the nonlinearity of transport maps, we introduce two variants of the proposed filter, the PSMF-L and PSMF-NL, which are based on linear and nonlinear maps, respectively. The PSMF is tested on various nonlinear state-space models and a nonlinear non-Gaussian target tracking model. The proposed linear PSMF-L outperforms all the reference models for medium-to-large numbers of particles, whilst the PSMF-NL shows better resilience to parameter changes.

1. Introduction

The state-space formulation for time-dependent models has been long used in various applications in science and engineering. For example, in target tracking, the state vector represents the kinematics of the target [1–3]; in weather prediction, it is related to temperature, pressure, humidity, etc.; in oceanography, it could refer to the spatial pattern of surface currents; and in economics, it concerns interest rates, inflation, etc. Nevertheless, filtering based on such formulation still faces numerous challenges, especially in nonlinear and high-dimensional situations. In the literature, various types of filters have been proposed, to cope with the various state estimation issues encountered by different models.

Indeed, in the case of linear Gaussian signal and observation models, the minimum mean squared error (MMSE) estimator underpins the Kalman filter (KF) [4]. However, despite its optimality in the aforementioned scenario, the KF is not applicable when either model is nonlinear. To overcome this, various extensions of the KF have been proposed in the literature, such as Extended

Kalman filter (EKF) [5] and Unscented Kalman filter (UKF) [6,7]. Even though both EKF and UKF constitute improvements on the classical KF, they still have some drawbacks. Among those, satisfactory performance can only be achieved when they are used in mild nonlinear conditions. The reason behind this is the model linearization in EKF and the unscented transform in UKF have a weak capability to approximate high-nonlinear dynamic behaviours.

To address solutions to the state estimation for highly nonlinear and non-Gaussian models, the particle filter (PF) [8–10] has been proposed and is one of the most common choices in the literature. The key idea behind this important filtering approach is to represent the posterior density function in terms of a set of random samples along with their associated weights and to compute estimations based on these weighted samples. In cases when the number of samples becomes large (i.e., the Monte Carlo (MC) simulation), the distribution of these samples corresponds to an equivalent representation to the functional description of the posterior probability density function (pdf). In other words, large number of samples help to represent posterior distribution perfectly and the PF estimates converge to the optimal Bayesian estimates. Despite its ability to model non-linear and non-Gaussian cases, the PF suffers from sample degeneracy especially in large-scale systems [11], where all but one particle weights tend to zero after a few iterations [8]. Thus the particles along with their corresponding weights cannot accurately represent the posterior distribution

* Corresponding author.
E-mail addresses: ju18422@bristol.ac.uk (P. Hao), karakuso@cardiff.ac.uk (O. Karakuş), alin.achim@bristol.ac.uk (A. Achim).

any longer. To alleviate the particle degeneracy, many variants of PF have been presented based on two strategies: (1) the use of effective importance proposals [8,12], such as auxiliary PF [13], unscented PF [14], and PF based on multiple importance sampling [15]; (2) employing resampling techniques [16,17], such as multinomial resampling [18], systematic resampling [16] or residual resampling [19]. Despite these improvements, in the highly nonlinear and high-dimensional cases, a large number of particles are still required due to the scale of the degeneracy problem and the computation load becomes prohibitive.

The Ensemble Kalman filter (EnKF) [20–22] has been proposed as another solution for the highly nonlinear cases. The EnKF also has a deterministic version which is the Ensemble Square Root filters (ESRF) [23]. In particular, for EnKF, the linear transformation is performed statistically by treating the observations as random variables. By contrast, in ESRF, analysis perturbations follow the KF analysis error covariance equation. Different from the PF, both EnKF and ESRF just require a small number of particles as they perform posterior estimation through linear particle movements and avoid the computation of importance sampling. However, despite these advantages over the PF, their performance in highly nonlinear environments is limited due to the intrinsic bias resulting from the low flexibility of their linear updates.

In an attempt to increase the flexibility of filtering techniques and address the drawbacks mentioned above, two kinds of nonlinear-transportation-based particle filters have been proposed. First, there are particle flow filters, which assume that the particle flow is embedded in a reproducing kernel Hilbert space and efficient solutions can be designed [24–26]. Although these approaches can provide flexible particle movements, they suffer from the slow transition from prior to posterior space. To improve transport efficiency, for the second method, optimal transport (OT) theory has also been utilised in conjunction with Bayesian methods to design one-step transportation methods. Optimal transport is an important mathematical theory, which was first introduced by Monge and then developed by Kantorovich [27]. Optimal transport maps define couplings that minimize a particular integrated transport cost corresponding to rearranging samples. Several filters based on optimal transport theory have been proposed. In [28], the ensemble transform particle filter (ETPF) has been introduced, whereby the optimal transport problem is resolved by linear programming. Despite its robust filtering performance, it suffers from a large computational load. In [29], two efficient generalizations of EnKF and ESRF have been presented, which are the stochastic map filter (SMF) and the deterministic transport map filter (DMF).For both of these approaches, the nonlinear movements of particles are captured by stochastic or deterministic maps based upon Knothe–Rosenblatt (KR) coupling which is also an optimal transport coupling [30]. Compared with other OT maps, the KR map has a triangular structure, so it is simple to invert, and computationally easier to parameterize and learn [27,31]. Benefiting from the flexibility of the nonlinear maps, the SMF has produced better performance for nonlinear models when compared to the EnKF which is based on linear transformation.

In recent years, motivated by the aforementioned drawbacks of PF and EnKF, three hybrid filters have been proposed and explored, namely the Gaussian mixture model-EnKF hybrid filter (GMM-EnKF) [32,33]; the hybrid of ETPF and ESRF (ETPF-ESRF) [28,34]; and the SIR-ESRF [35] via combining the standard PF and ESRF with a mean-preserving random orthogonal resampling. For all the aforementioned hybrid filters, the likelihood function is first separated into two parts, then, EnKF/ESRF and PF-based filters are applied sequentially to assimilate each likelihood part. As a result, the hybrid filters can yield more precise non-Gaussian approximations than PF since they alleviate the particle degeneracy issue through the use of EnKF/ESRF. However, these filters also have some drawbacks. In GMM-EnKF, since the EnKF is implemented in the first stage, it is not suitable for moderately non-Gaussian models where prior distributions are non-Gaussian and posterior distributions are close to Gaussian. ETPF-ESRF suffers from a large computational load due to the ETPF component. SIR-ESRF provides an efficient solution to the moderately non-Gaussian models and the mean-preserving random orthogonal transformation in the ESRF update stage produces the Gaussian approximation for the posterior distribution. However, in cases when the posterior is not close to a Gaussian form, this resampling technique causes large sampling errors.

In this paper, we propose a new nonlinear filter by combining ideas from the PF and the SMF. In particular, the main contributions of this work consist in:

1. We propose a hybrid particle-stochastic map filter, referred to as PSMF, which combines the particle filter and the stochastic map filter. In order to build the new hybrid filtering approach, the likelihood function is separated into two parts, with the PF and SMF being then employed sequentially. In the proposed approach, we utilise SMF and not DMF since the former only requires the optimisation of a convex function.
2. To address the particle degeneracy issue, we employ the systematic resampling method [16,17] and a subsequent smoothing step by adding a low-variance Gaussian noise. Compared with the ETPF, the computation load of systematic resampling is relatively low and proportional to the number of particles. Also, it is not based on Gaussian approximation, hence when the posterior distribution is not close to Gaussian, the adopted resampling technique is still efficient.
3. To evaluate the influence of the nonlinearity of the transport maps on hybrid filters, we introduce PSMF-L and PSMF-NL, which are based on linear and nonlinear transport maps, respectively. These two variants were compared in the experimental analysis, and their advantages and disadvantages are analysed in details.

The rest of the paper is organised as follows: we begin in Section 2 with a review of the theoretical preliminaries in the form of state-space models, particle filters, and stochastic map filter. In Section 3, we present the details of the proposed PSMF algorithm. The experimental analysis is performed in Section 4, whilst Section 5 concludes the paper with a summary.

2. Background on particle filtering

2.1. State-space model

A state-space model can be described as a probabilistic graphical model, which defines the relationship between the hidden state variables and the measurements. It includes two equations, which are for states (signals) and their measurements. The signal model describes the state changes over time, whilst the measurement model explains the relationship between the states and measurements. A generic state-space model is expressed as

$$X_k = f(X_{k-1}, U_{k-1}) \tag{1}$$

$$Z_k = h(X_k, W_k) \tag{2}$$

where X_k refers to the hidden state, Z_k is the observation, the random model error is denoted by U_k, and W_k refers to the observation noise at time k. The random model and observation noise errors are independent of each other and the states. Figure 1 depicts the general conditionally independent structure of state-space models, which shows the independence of observations conditioned on the state.
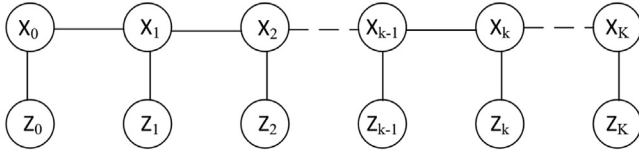
Fig. 1. State-space model.

2.2. Sequential Monte–Carlo methods

Particle filters, also referred to as Sequential Monte Carlo (SMC) algorithms are a class of Monte Carlo algorithms used to solve some filtering problems arising in signal processing and in Bayesian statistical inference. They represent a popular solution to estimate the expectations of the posterior distribution by representing the density with a large number of weighted samples. The key step of the PF algorithm is to design proper importance proposals for the posterior distribution and use the importance sampling method to calculate its mean. Thus, estimating the expectation of the posterior distribution becomes a tractable problem through the samples and their corresponding weights. At any time k, the posterior distribution can be expressed as

$$p(X_{0:k}|Z_{1:k}) \propto p(Z_k|X_k)p(X_k|X_{k-1})p(X_{0:k-1}|Z_{1:k-1}). \qquad (3)$$

Generally, the posterior distribution has a complex form, and sampling from it cannot be done. Hence, we usually obtain the sampled particles of posterior distributions using the importance sampling methodology. Let the importance proposal be $q(X_{0:k}|Z_{1:k})$, then we take the sequential form of

$$q(X_{0:k}|Z_{1:k}) = q(X_k|X_{0:k-1}, Z_{1:k})q(X_{0:k-1}|Z_{1:k-1}) \qquad (4)$$

where we firstly sample from the importance proposal and obtain the particles $X_{0:k}^i, i = 1 \ldots N$. Then, the corresponding weights w_k^i can be expressed as

$$w_k^i \propto \frac{p(X_{0:k}^i|Z_{1:k})}{q(X_{0:k}^i|Z_{1:k})}. \qquad (5)$$

Hence, the posterior density is approximated as

$$p(X_{0:k}|Z_{1:k}) \approx \sum_{i=1}^{N} w_k^i \delta(X_{0:k} - X_{0:k}^i). \qquad (6)$$

For the standard particle filter formulation, the importance proposal is chosen as

$$q(X_k^i|X_{k-1}^i, Z_k) = p(X_k^i|X_{k-1}^i) \qquad (7)$$

which relies only on the prior information and leads to the weight updating equation as

$$w_k^i \propto w_{k-1}^i p(Z_k|X_k^i). \qquad (8)$$

Degeneracy problem As mentioned in the above sections, a common problem in particle filtering applications is the degeneracy phenomenon, where a large part of the particles will have negligible weights after a few iterations. It has been proved that the variance of the importance weights can only increase over time [8], and thus it is impossible to avoid the degeneracy phenomenon. Possibly, the most effective way to reduce the effects of this drawback is to design an efficient importance proposal [9]. Also, while PF will always degenerate in the long term, this problem can be alleviated by resampling techniques [16,17], which replicate the particles with large weights and eliminate the ones with small weights.

3. Stochastic map filtering

In this section, the general formulation of the SMF is introduced. We first provide some brief, necessary background on OT. Then, we explain the Knothe–Rosenblatt rearrangement, which is an optimal transport map designing procedure and constitutes the basis of the SMF approach. The cost function for the parameter optimisation of the KR rearrangement is introduced subsequently. Finally, the assimilation step of the SMF concludes this subsection.

3.1. Optimal transport fundamentals

The optimal transport [27,36] is a historical mathematical theory, which exploits the most efficient way to move one distribution to another. It was first presented by French mathematician Gaspard Monge in 1781. Specifically, Monge's optimal transportation problem pursues a transport map g which pushes one probability distribution $p(X), X \in \Omega_0$ to another $q(Y), Y \in \Omega_1$, i.e.,

$$\int_{X:g(X)\in B} P(X)dX = \int_B q(Y)dY \qquad (9)$$

where B is an arbitrary area in the domain Ω_1. In general, there exists an infinite number of maps that solve Eq. (9). One way to select a map is to define a particular cost function and solve the optimal transport problem. For example, g can be restricted by minimising the p-Wasserstein metric below

$$M(p, q) = \min_{g \in MP} \left(\int_{\Omega_0} c(X, g(X))p(X)dX \right)^{\frac{1}{p}}, \qquad (10)$$

where MP denotes measure preserving maps meeting Eq. (9). $c(X, g(X)) = |X - g(X)|^p, \ p \geq 1$ is a cost function, and denotes the transport cost from X to $g(X)$.

In 1942, Monge's transport problem was generalised by Leonid V. Kantorovich, whose formulation pursues a transportation plan γ. One can regard γ as the joint distribution which has marginals $p(X)$ and $q(Y)$, i.e.,

$$\begin{aligned} \gamma(\Omega_0 \times Y) &= q(Y) \\ \gamma(X \times \Omega_1) &= p(X). \end{aligned} \qquad (11)$$

Its corresponding objective function can be expressed as

$$K(p, q) = \min_{\gamma \in MP} \left(\int_{\Omega_0 \times \Omega_1} c(X, Y)d\gamma(X, Y) \right)^{\frac{1}{p}}, \qquad$$

where MP meets Eq. (11), and $c(X, Y) = |X - Y|^p, \ p \geq 1$. We should note that the optimization problem over the joint distribution is over a different space than the MP used in Eq. (10). Specifically, for Monge's problem, a transport map assigns one element of X to exactly one of Y. By contrast, Kantorovich's problem is more general and supports one-to-many movements.

In recent years, the regularisation of many optimal transport maps/plans has been presented and applied to filtering problems, such as ETPF [28], the mapping particle filter [25], and filtering based on the Knothe–Rosenblatt (KR) rearrangement [29].

3.2. Knothe–Rosenblatt rearrangement

Given any pair of positive densities, there exists a unique monotone triangular transport map, which defines a deterministic coupling between two distributions and is called the KR rearrangement [36]. This strategy belongs to Monge's problem, i.e., a one-to-one transport map. Due to its ability to create connections between two probability densities, it has been widely used in Bayesian inference, importance sampling, SMC, etc. [37–39].

A parametric strategy for the KR rearrangement is presented in Moselhy and Marzouk [37]. Assume $X = [x_1, x_2, \ldots, x_n]^T$ and

$Y = [y_1, y_2, \ldots, y_n]^T$ are n-dimensional variables with distributions $p(X)$ and $q(Y)$, respectively. Then, a standard triangular transport map can be expressed as

$$Y = S(X) = \begin{bmatrix} S^1(x_1) \\ S^2(x_1, x_2) \\ \ldots \\ S^n(x_1, x_2, \ldots, x_n) \end{bmatrix} \qquad (12)$$

where the map S transports the distribution $p(X)$ to $q(Y)$ and each component of S is monotone with respect to its last input. The crucial property of the KR rearrangement for its application to the SMC algorithm is that it provides an implicit characterization of the marginal conditional distributions. In the example above, S^1 transports $p(x_1)$ to $q(y_1)$, S^2 transports $p(x_2|x_1)$ to $q(y_2|y_1)$ and so on. This conditional distribution transformation property is used in the design of the SMF algorithm.

The parameterisation of KR rearrangement is an important part of the design of transport maps. One way to parameterise each component of the map S is via multivariate polynomials which could either involve Hermite or Legendre polynomials [31]. Spantini et al. [29] also provides a specific computational parameterisation of triangular maps, which is based on radial basis functions (RBFs), and we adopt the same form in this work.

3.3. Constructing KR rearrangement from samples

Following the definition of the triangular transport maps above, we now explain the cost function definition based on the Kullback–Leibler (KL) divergence. We aim to optimise the parameters of the transport map polynomials which connect an arbitrary distribution $p(X)$ to a reference distribution $q(Y)$. Following Eq. (9), since S is a monotone and differentiable transformation, we have

$$\hat{p}(X) = q(S(X))|\det(D_X S)|, \qquad (13)$$

where $\hat{p}(X)$ is an approximation of $p(X)$. The map S can be obtained by minimizing the difference between $\hat{p}(X)$ and $p(X)$, and the KL divergence can be used to measure this. It can be expressed as

$$\begin{aligned} D_{KL}(p|\hat{p}) &= \mathbb{E}\left[\log \left(\frac{p(X)}{\hat{p}(X)} \right) \right] \\ &= \mathbb{E}[\log(p(X)) - \log(q(S(X))) - \log|\det(D_X S)|]. \end{aligned} \qquad (14)$$

Then, the transport map can be expressed in terms of the KL divergence as

$$\hat{S} = \arg\min_{S \in \mathcal{H}} \mathbb{E}\left[-\log(q(S(X))) - \log|\det(D_X S)| \right], \qquad (15)$$

where $\mathcal{H}$ is a function space for map S. The unknown term $\mathbb{E}[\log(p(X))]$ is neglected in the objective function above because it is not related to S. We should note that in the above cost function, there is no term related to transport cost like Eq. (10), because the uniqueness of the transport map is guaranteed by the triangular structure and monotonicity of KR rearrangement [37]. Assume that we have N samples X^i, $i = 1 \ldots N$, from p(X), then the cost function based on the discrete samples is expressed as

$$\hat{S} = \arg\min_{S \in \mathcal{H}} \frac{1}{N} \sum_{i=1}^{N} \left[-\log q(S(X^i)) - \log|\det(D_X S(X^i))| \right] \qquad (16)$$

Newton's method provides an efficient solution to the minimisation of the cost function in (16) and has been widely used in the literature [29].

3.4. Data assimilation

In this section, we introduce the structure of the SMF. For a filtering problem, at any time k, we initially have samples $X^i_{k-1|k-1}$, $i = 1, \ldots, N$, from $p(X_{k-1}|Z_{1:k-1})$. By sampling

$p(X_k|X^i_{k-1|k-1})$, we have $X^i_{k|k-1}$, $i = 1, \ldots, N$. Then by sampling $p(Z_k|X^i_{k|k-1})$, $Z^i_{k|k-1}$, $i = 1, \ldots, N$ are obtained. Then we have $\left[X^{i\,T}_{k|k-1}, Z^{i\,T}_{k|k-1} \right]^T$, $i = 1, \ldots, N$ which are samples from the joint density $p(X_k, Z_k|Z_{1:k-1})$. Following [29], we define the function $S^{\mathcal{X}} : R^n \times R^m \to R^n$, which is a part of the KR rearrangement $S : R^n \times R^m \to R^n \times R^m$, which transports the distribution $p(X_k, Z_k|Z_{1:k-1})$ to a normal distribution. The resulting map is continuous and can be defined as

$$S^{\mathcal{X}}(Z_k, X_k) \sim \mathcal{N}(0, \mathcal{I}_n) \qquad (17)$$

From [29], we have that the analysis map T which transports the joint distribution to the posterior can be expressed as

$$T := S^{\mathcal{X}}(Z_k^*, \sim)^{-1} \circ S^{\mathcal{X}}(Z_k, X_k) \qquad (18)$$

where Z_k^* is a fixed observation at time k, $S^{\mathcal{X}}(Z_k^*, \sim)^{-1}$ denotes the inverse function of the map $S^{\mathcal{X}}(Z_k^*, \sim)$, and $\circ$ denotes the composition of the two maps. In the above structure, $S^{\mathcal{X}}$ transforms the samples from the joint distribution to the standard normal which is then pushed by $S^{\mathcal{X}}(Z_k^*, \sim)^{-1}$ to the posterior. In principle, we can directly employ $S^{\mathcal{X}}(Z_k^*, \sim)^{-1}$ to produce posterior approximations by pushing forward samples from the standard normal density. However, from Spantini et al. [29], transporting samples from $p(X_k, Z_k|Z_{1:k-1})$ through T yields more accurate results which is attributed to the cancellation of errors in the composition of $S^{\hat{\mathcal{X}}}$ and its inverse.

In practice, we can obtain the estimator of T using the samples, and it can be expressed as

$$\hat{T} := S^{\hat{\mathcal{X}}}(Z_k^*, \sim)^{-1} \circ S^{\hat{\mathcal{X}}}(Z_k, X_k) \qquad (19)$$

where $S^{\hat{\mathcal{X}}}$ denotes the estimated map of $S^{\mathcal{X}}$.

In general, the process of stochastic map filtering can be divided into two steps, i.e., the forecast step and the analysis step. First, we get samples of $p(X_k, Z_k)$ (forecasting). Second, the samples of the estimated posterior distribution can be obtained by transforming the joint prior samples through the map $\hat{T}$ (analysis).

4. Stochastic map-augmented particle filtering

In this section, we present the proposed hybrid filter that combines PF with the SMF. At any time k, we have samples $(X^i_{k-1}, i = 1, \ldots, N)$ subject to $p(X_{k-1}|Z^*_{1:k-1})$. Then, forecast samples $(X^i_k, i = 1, \ldots, N)$ from $p(X_k|Z^*_{1:k-1})$ can be obtained by sampling from $p(X_k|X^i_{k-1})$. After the prior samples are obtained, the likelihood function is separated as

$$p(Z_k^*|X_k) = p(Z_k^*|X_k)^{\alpha} p(Z_k^*|X_k)^{1-\alpha} \qquad (20)$$

where $0 < \alpha < 1$ is the *splitting factor*, $p(Z_k^*|X_k)^{\alpha}$ is assimilated by PF and $p(Z_k^*|X_k)^{1-\alpha}$ is assimilated by SMF. For α values of 0 and 1, the hybrid filter boils down to the SMF and the PF, respectively.

When α is small, PSMF is close to the SMF. In this case, the sample degeneracy can be alleviated effectively, but the estimation error increases due to the limited flexibility of the maps. We should note that theoretically, the degree of non-linearity of a map can be arbitrarily high and there is no limit to its flexibility. However, with a certain number of particles, highly non-linear maps suffer from larger map variances and may produce worse estimations than more linear maps. So, the flexibility of nonlinear maps should be controlled according to the ensemble size. Specially, for a small number of particles, less non-linear maps, should be used. By contrast, a larger number of particles can restrict the variance of a highly non-linear map [29].

By contrast, for large α, PSMF approaches closely the PF. Then, the error from sample degeneracy increases. In this case, as $1 - \alpha$ tends to zero, $p(Z_k^*|X_k)^{1-\alpha}$ just reshapes the distribution slightly.

So, there is no need for a highly flexible transport to complete the second assimilation step, and the SMF-part does not cause a rough estimate. With the adaptability mentioned above, the proposed hybrid filter achieves a trade-off between the PF and the SMF.

Before we implement the assimilation step, α needs to be chosen. For a given θ, we choose α as the largest value which satisfies the condition

$$ESS \geqq \theta N \tag{21}$$

where ESS is the effective sample size and can be approximated by the sample weights. For a given α, based on (8), we have the weight update equation which can be expressed as

$$w_k^i \propto w_{k-1}^i p(Z_k^*|X_k^i)^\alpha. \tag{22}$$

Since a resampling step is taken at every iteration, we have $w_{k-1}^i = 1/N$ and the normalized particle weights can be written as

$$w_k^i = \frac{p(Z_k^*|X_k^i)^\alpha}{\sum_{j=1}^N p(Z_k^*|X_k^j)^\alpha}. \tag{23}$$

Then, the ESS can be approximated by

$$ESS = \frac{1}{\sum_i (w_k^i)^2} \tag{24}$$

From [33], searching the value of $0 \leqslant \alpha \leqslant 1$ satisfying the condition in (21) can be achieved by a root finding method. For large θ, the searched α is small, and then the hybrid filter is closer to SMF. For a smaller θ, the hybrid filter moves towards the PF. The optimal values of θ depend on the nonlinear models. Currently, there is no automatic way to select parameter θ, and it needs to be selected manually based on the models. After θ is selected and α is calculated, the PF is implemented to update $p(Z_k^*|X_k)^\alpha$. Based on (23), we obtain the normalized particle weights corresponding to the selected α.

To alleviate the particle degeneracy problem, we utilise a systematic resampling followed by a smoothing step. Different strategies for the alleviation of particle degeneracy [33–35] have been proposed in the literature. For example, EnKF-GMM avoids the degeneracy by sampling from the posterior density approximated by a Gaussian Mixture; the ETPF-ESRF filter employs the optimal transport strategy to implement the resampling; a mean-preserving random orthogonal transformation is used in SIR-ESRF to break the degeneracy. Although the performance of systematic resampling is proved in previous work, one did not pay enough attention to its embedding in hybrid filters, because the discontinuity update influences its application in the high-dimensional scenario. In this work, we only consider the low-dimension cases, but with the developments of the local PF [24,40] and its combination with the EnKF [41], we can reasonably consider applying our hybrid filter to high-dimension models. This however is out of the scope of this paper.

The proposed resampling process [17] starts by generating N ordered numbers $u_i, i = 1 : N$ by

$$u_i = \frac{(i-1) + u}{N}, u \sim U[0, 1) \tag{25}$$

which are used to select X_k^i according to the multinomial distribution. That is

$$X_k^i = X(F^{-1}(u_i)), \ s.t. u_i \in \left[\sum_{s=1}^{i-1} w_k^s, \sum_{s=1}^i w_k^s \right] \tag{26}$$

where function $X(i) = X_k^i$. F denotes the cumulative probability distribution of the normalised particle weights, i.e., $F(i) = \sum_{s=1}^{i-1} w_k^s$. F^{-1} denotes the inverse of F. After resampling, all the particle weights become equal, $w_k^i = 1/N$.

After resampling, to increase the diversity of particles, we add a small Gaussian noise to smooth the samples, which is helpful to decrease the variance of transport maps in the SMF assimilation stage.

$$X_k^i = \overline{X}_k + \zeta(X_k^i - \overline{X}_k) + \beta * v_i, v_i \sim N(0, var(X_k)) \tag{27}$$

where $0 < \beta < 1$ is the smoothing factor and $var(X_k)$ is the variance of the resampled samples. $\zeta = \sqrt{1 - \beta^2}$ is the shrinking factor, which is employed to remove the excess variance caused by the added noise [32].

Next, $(p(Z_k^*|X_k)^{1-\alpha}$ is assimilated by the SMF. First, to obtain the samples $[Z_k^i, X_k^i]$ from the joint distribution, we sample the likelihood $p(Z_k|X_k^i)^{1-\alpha}, i = 1, \ldots, N$. In this work, we only consider Gaussian and Gaussian Mixture likelihood. For Gaussian cases, assume $p(Z_k|X_k^i) = \mathcal{N}(f(X_k^i), \Sigma)$, i.e.,

$$p(Z_k|X_k^i) = \frac{1}{(2\pi)^{\frac{m}{2}} |\Sigma|^{\frac{1}{2}}} \mathbf{exp}\left[-\frac{1}{2}(Z_k - f(X_k^i))^T \Sigma^{-1}(Z_k - f(X_k^i)) \right]$$

where Σ is the covariance matrix. Then

$$p(Z_k|X_k^i)^{1-\alpha}$$
$$= \left[\frac{1}{(2\pi)^{\frac{m}{2}} |\Sigma|^{\frac{1}{2}}} \right]^{1-\alpha} \mathbf{exp}\left[-\frac{1}{2}(Z_k - f(X_k^i))^T \left(\frac{\Sigma}{1-\alpha} \right)^{-1}(Z_k - f(X_k^i)) \right]$$
$$\propto \frac{1}{(2\pi)^{\frac{m}{2}} \left| \frac{\Sigma}{1-\alpha} \right|^{\frac{1}{2}}} \mathbf{exp}\left[-\frac{1}{2}(Z_k - f(X_k^i))^T \left(\frac{\Sigma}{1-\alpha} \right)^{-1}(Z_k - f(X_k^i)) \right]$$
$$= \mathcal{N}\left(f(X_k^i), \frac{\Sigma}{1-\alpha} \right).$$

Hence, we can sample $\mathcal{N}(f(X_k^i), \frac{\Sigma}{1-\alpha})$ to obtain Z_k^i. For Gaussian mixture cases, the likelihood can be expressed as $p(Z_k|X_k^i) = \sum_{j=1}^M \mathcal{P}_j \mathcal{N}(f(X_k^i), \Sigma_j)$ where M is the number of mixtures. $\mathcal{P}_j$ and Σ_j are the mixing factor and the covariance matrix of the ith Gaussian component, respectively. Similarly, we can sample $\sum_{j=1}^M \mathcal{P}_j \mathcal{N}\left(f(X_k^i), \frac{\Sigma_j}{1-\alpha} \right)$ to obtain Z_k^i.

Then, the map $\hat{S}^{\mathcal{X}}$ can be designed from the samples $[Z_k^i, X_k^i]$ according to (17) and the transport map $\hat{T}$ can be designed by (19).

In the last step, the samples $[Z_k^i, X_k^i]$ are transported with the transport map $\hat{T}$ as

$$X_k^i = \hat{T}(Z_k^i, X_k^i) \tag{28}$$

In our work, thanks to the generalisation of the EnKF, the transport map $\hat{T}$ can be designed to be both linear and nonlinear which can then respectively be called PSMF-L and PSMF-NL. As also noted in Spantini et al. [29], the nonlinear map achieves a better performance compared to the linear map, but suffers from larger map variances. The overall procedure of the proposed PSMF algorithm is summarised in Algorithm 1.

5. Simulation results

5.1. State-space models

In this section, we introduce the low-dimensional nonlinear state-space models used in the experimental analysis. There are four benchmark models considered, i.e., (1) the univariate non-stationary growth model [42], (2) Henon map [35], (3) Lorenz-63 model [29], and (4) a target tracking model also used in Henke et al. [43],44]. The first model is one-dimensional, the Henon map is a two-dimensional model, whilst Lorenz-63 and the target tracking models are of dimensions 3 and 4, respectively. The details of these state-space models are presented in the sequel.

Algorithm 1: The PSMF algorithm

procedure PSMF $(Z_k^*, p(X_k|X_{k-1}), p(Z_k|X_k), \theta, X_{k-1}^i)$;**for** $i \leftarrow 1$
to N **do**
  $X_k^i \leftarrow$ sample from $p(X_k|X_{k-1}^i)$;

$\alpha \leftarrow$ root-finding method;**if** $\alpha > 0$ **then**
  for $i \leftarrow 1$ *to N* **do**
    $w_k^i \leftarrow$ particle weights by (23);
  for $i \leftarrow 1$ *to N* **do**
    $X_k^i \leftarrow$ systematic resampling by (25) and (26);
  for $i \leftarrow 1$ *to N* **do**
    $X_k^i \leftarrow$ smoothing by (27);

if $\alpha < 1$ **then**
  for $i \leftarrow 1$ *to N* **do**
    $[Z_k^i, X_k^i] \leftarrow$ sample from $p(Z_k|X_k)^{1-\alpha}$;
  $\hat{S}^{\mathcal{X}} \leftarrow$ estimator of $S^{\mathcal{X}}$ from $[Z_k^i, X_k^i]$ by (16). $S^{\mathcal{X}}$ is defined
  by (17);**for** $i \leftarrow 1$ *to N* **do**
    $X_k^i \leftarrow \hat{S}^{\mathcal{X}}(Z_k^i, X_k^i)$;
    $X_k^i \leftarrow \hat{S}^{\mathcal{X}}(Z_k^*, X_k^i)^{-1}$;

return X_k^i

5.1.1. Univariate non-stationary growth (UNGM) model

The UNGM model can be expressed as

$$x_k = \alpha x_{k-1} + \beta \frac{x_{k-1}}{1 + x_{k-1}^2} + \gamma \cos(1.2(k-1)) + u_k \tag{29}$$

$$y_{k+1} = x_{k+1} + w_{k+1}$$

where $\alpha = 0.5, \beta = 25, \gamma = 8$ and $u_k \sim \mathcal{N}(0, 1), w_k \sim \mathcal{N}(0, 2.5)$ are zero-mean Gaussian noise. In our experiments, we set the total number of steps to 100 and repeat the simulations 100 times. The initial state follows $N(20, 1^2)$.

5.1.2. Henon map

Different from the other models, for the Henon map, the number of steps is set to 1. This helps to focus on a single Bayesian assimilation update. Hence, the complication related to varying non-Gaussianity along cycled steps can be eliminated.

From [35], the prior is the joint density of U and V whose samples are obtained by pushing samples of normal variables U_0 and V_0 forward through the Henon map, i.e.,

$$U = 1 - 1.4U_0^2 + V_0,$$
$$V = 0.3U_0 \tag{30}$$

where $U_0 \sim \mathcal{N}(0, 1^2), V_0 \sim \mathcal{N}(0, 0.1^2)$. The true values of U and V are set to -4 and 0.6. 1000 Monte Carlo runs were simulated in the experiments.

5.1.3. Lorenz-63

The Lorenz-63 Model is a 3-dimensional nonlinear state-space model, and its state vector is represented by $X_t = (X_1(t), X_2(t), X_3(t))$. The dynamics of the model can be described by

$$\frac{dX_1}{dt} = -\sigma X_1 + \sigma X_2$$

$$\frac{dX_2}{dt} = -X_1 X_3 + \rho X_1 - X_2 \tag{31}$$

$$\frac{dX_3}{dt} = X_1 X_2 + \beta X_3$$

where β, ρ, σ are fixed parameters, which are set to $\beta = 8/3, \rho = 28, \sigma = 10$. The ODE system above is integrated by a fourth-order explicit Runge–Kutta method and a constant stepsize of $\triangle t = 0.05$. Also, the process noise follows a Gaussian distribution $\epsilon_{\triangle t} \sim \mathcal{N}(0, 10^{-4}I_3)$.

For the observation model, we assume the states are measured every $\triangle t_{obs} = 0.5$ time units, then the measurement model can be represented as

$$Y_k = Z_k + W_k \tag{32}$$

where $Z_k = Z(k\triangle t_{obs})$ and $W_k \sim \mathcal{N}(\theta^2 I_d)$ is zero-mean white Gaussian observation noise. In our experiment, the states are fully measured, i.e., $d = 3$, and the observation noise variance is set to $\theta^2 = 4$.

For the initialisation, the true state is set by sampling from $Z_0 \sim \mathcal{N}(0, I_d)$. The true model runs 6000-time steps to generate the states, and the observations are produced by the measurement models. The initial ensemble is produced by a spin-up stage. N samples are obtained from the initial condition $\mathcal{N}(0, I_d)$ and then the stochastic EnKF is run for the first 2000 steps. The produced particles are used as initial samples. In the last 4000 steps, several filters are run. Only the filter results for the last 2000 steps are used for the analysis of the filter quality, and the middle 2000 steps are used for eliminating the influence of switching between EnKF and other filters at the end of the 2000th iteration.

5.1.4. Target tracking with heavy-tailed measurement noise

In this subsection, we describe a nonlinear target tracking model [43,44]. As it is shown below, the state of the target X is a four-dimensional vector, where P_{east} and P_{north} are the map coordinates of the object, V_{abs} is the value of velocity, and φ is the direction of movement.

$$X = \begin{bmatrix} P_{east} \\ P_{north} \\ V_{abs} \\ \varphi \end{bmatrix} \tag{33}$$

The dynamics of the target tracking model can be described by

$$f(X) = \begin{bmatrix} P_{east} + \triangle t_{obs} * cos(\varphi) * V_{abs} \\ P_{north} + \triangle t_{obs} * sin(\varphi) * V_{abs} \\ V_{abs} \\ \varphi \end{bmatrix} + \epsilon \tag{34}$$

where $\triangle t_{obs}$ is the time interval from current state to the next state, and ϵ is the heavy-tailed process noise. Different from Henke et al. [43,44], in our experiment, the heavy-tailed noise is approximated by a Gaussian mixture as is used in various other target tracking models [2,3], and can be expressed as

$$\epsilon \sim \begin{cases} N(0, Q) & \text{with probability} \quad 0.85 \\ N(0, \eta Q) & \text{with probability} \quad 0.15 \end{cases} \tag{35}$$

where Q is the covariance matrix, and η is a diagonal matrix used to enlarge Q. Then, the measurement equation is given by

$$Z = HX + W \tag{36}$$

where $W \sim \mathcal{N}(0, R)$ is the measurement noise with covariance R and

$$H = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \tag{37}$$

is the transformation matrix. Following the measurement model, at every time instant, the position coordinates of the target, P_{east} and P_{north}, are observed. For the complete definition, we also need to specify the observation interval $\triangle t_{obs}$, the values for the initial state X_0 with the covariance matrix C_0, the processing noise covariance Q, and the measurement noise covariance R. In the
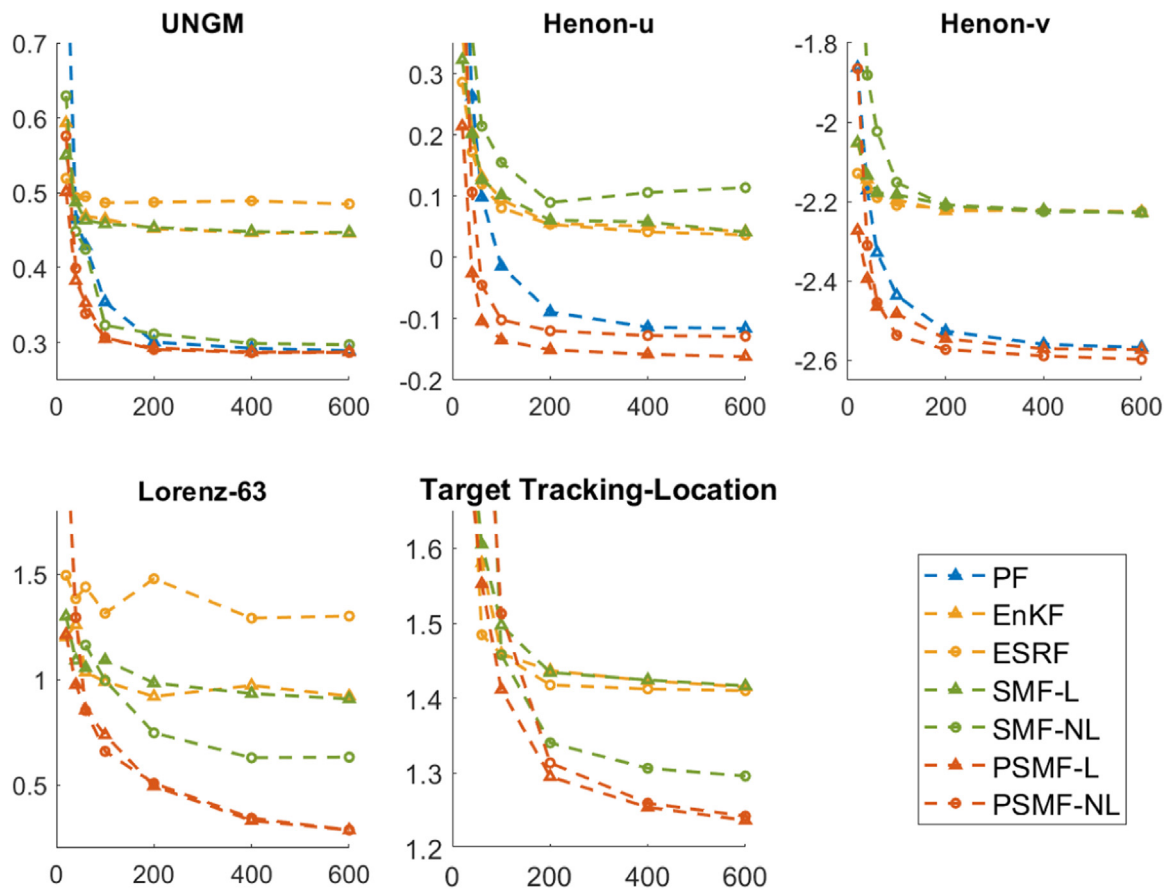
Fig. 2. The comparison of the proposed PSMF to the traditional filters under different number of particles. The horizontal axis represents the number of particles, and the vertical axis represents the log (RMSE).
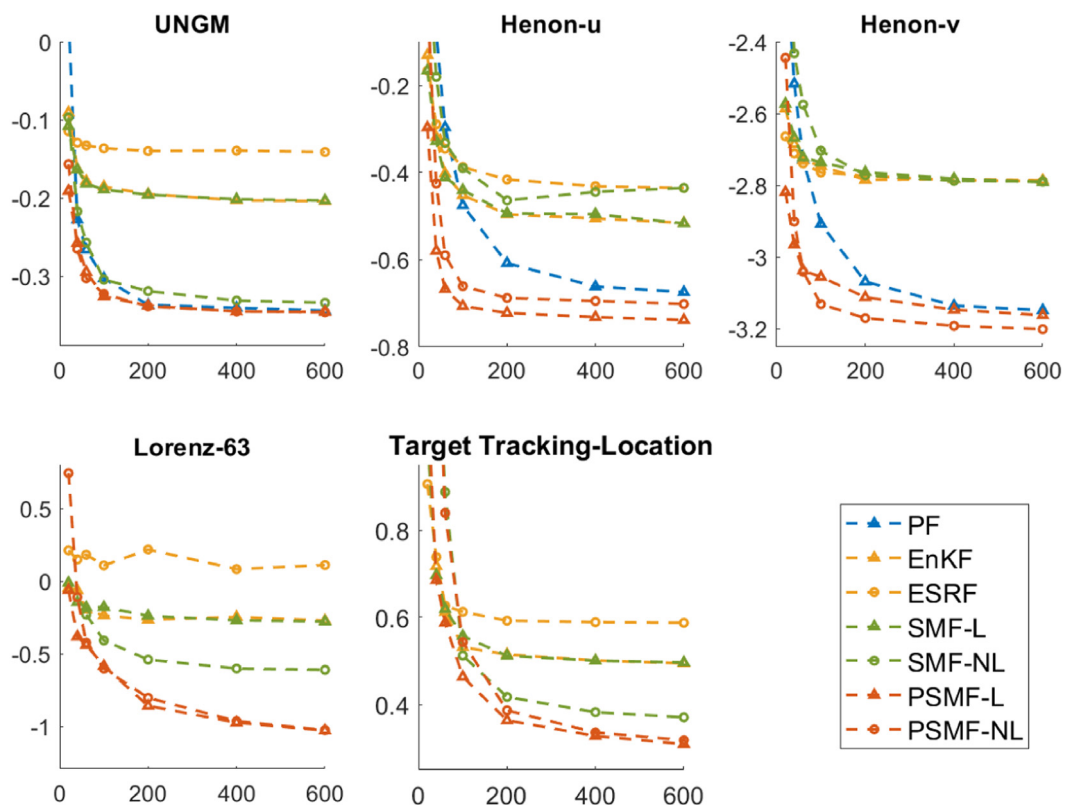

Fig. 3. The comparison of the proposed PSMF to the traditional filters under the different number of particles. The horizontal axis represents the number of particles, and the vertical axis represents the log (CRPS).
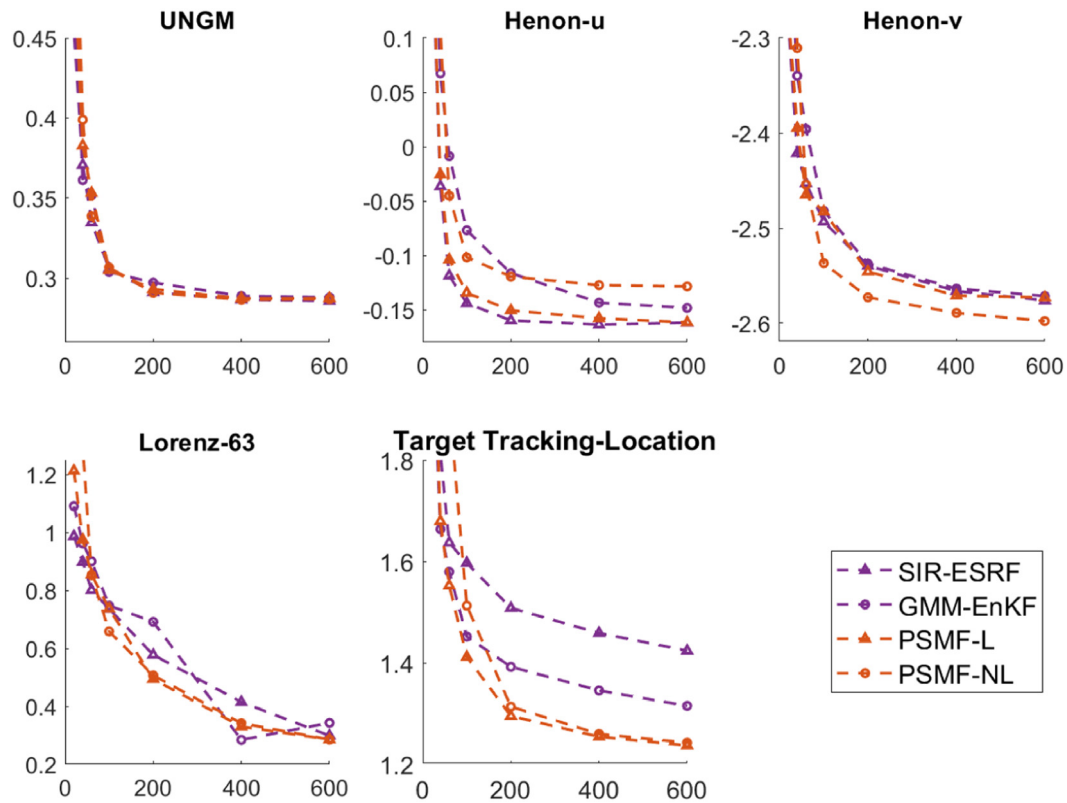
Fig. 4. The comparison of the PSMF to the other hybrid filters under the different number of particles. The horizontal axis represents the number of particles, and the vertical axis represents log(RMSE).
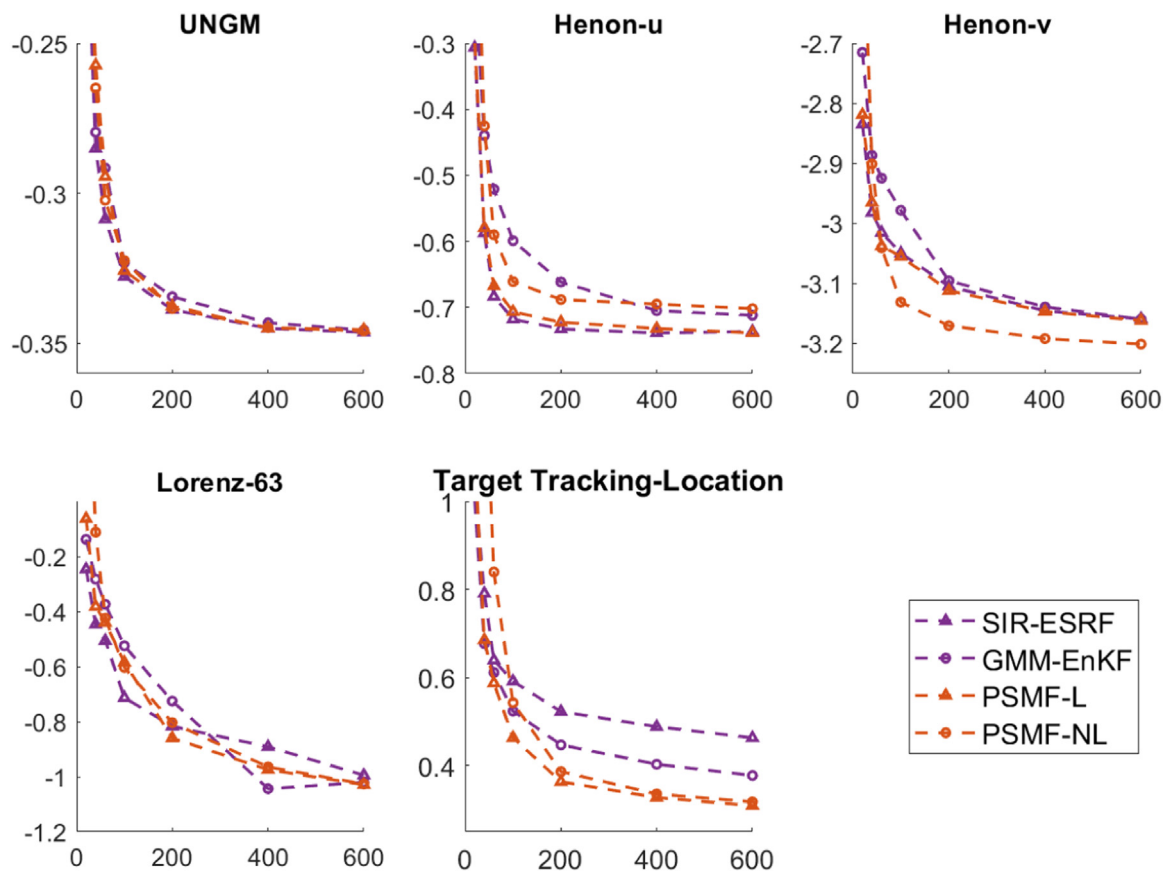

Fig. 5. The comparison of the proposed PSMF to the traditional filters under the different number of particles. The horizontal axis represents the number of particles, and the vertical axis represents log(CRPS).
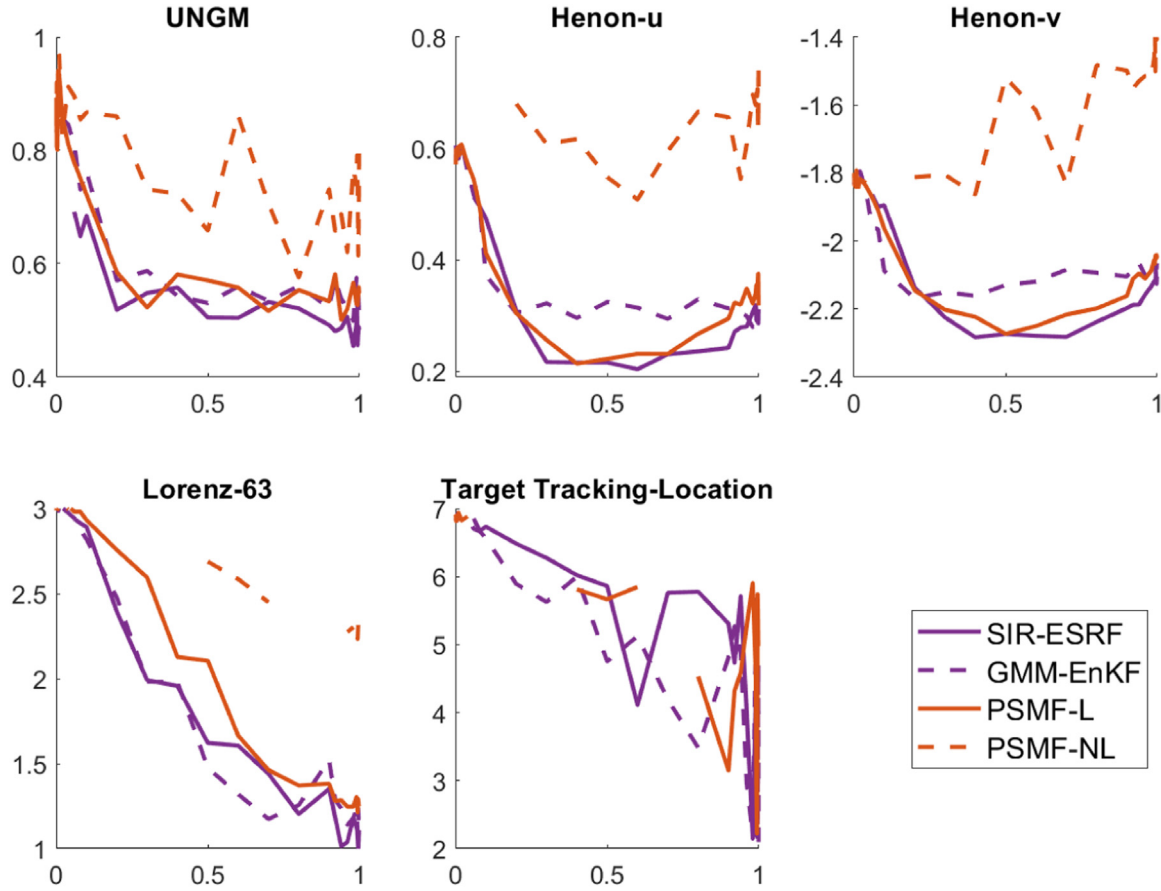
Fig. 6. The comparison among hybrid filters with different values of parameter θ under different models. The ensemble size N is set to 20. Horizontal axis represents the value of θ, and vertical axis represents the estimation log(RMSE).

experiment, without loss of generality, the target starts to move from the initial point and the real initial state X_0 is set to $\begin{bmatrix} 0 \text{ m} & 0 \text{ m} & 30 \text{ m/s} & 0 \text{ rad} \end{bmatrix}$, where the initial speed of 30 m/s is a reasonable value for a vehicle. For simplicity, we assume the initial state subject to a Gaussian distribution with the covariance C_0 where the deviation about the position, velocity, and direction is set to $\begin{bmatrix} 10 \text{ m} & 10 \text{ m} & 3 \text{ m/s} & \pi/10 \text{ rad} \end{bmatrix}$, so we have

$$C_0 = diag \begin{bmatrix} 100 \text{ m}^2 & 100 \text{ m}^2 & 9(\text{m/s})^2 & \pi^2/100 \text{ rad}^2 \end{bmatrix}.$$

In our experiments, we set

$$\triangle t_{obs} = 1,$$
$$Q = diag \begin{bmatrix} 0.01 \text{ m}^2 & 0.01 \text{ m}^2 & 0.01(\text{m/s})^2 & (\pi/90)^2 \text{ rad}^2 \end{bmatrix},$$
$$\eta = diag \begin{bmatrix} 100 & 100 & 100 & 900 \end{bmatrix},$$
$$R = diag \begin{bmatrix} 9m^2 & 9m^2 \end{bmatrix}.$$

The whole tracking process lasts for 120 s and 50 Monte Carlo runs were simulated.

5.2. Benchmark filters

In our experiments, the proposed hybrid PF-SMF filter is compared to 6 state-of-the-art filters, including the classic PF, EnKF, ESRF, SMF and two-hybrid filters of SIR-ESRF and GMM-EnKF. In this section, we introduce the configuration of these filters.

For PSMF-L and PSMF-NL, we set both the smoothing parameter β to 0.2. Following [29], linear and nonlinear transport maps take the "linear" and "linear + 2 RBFs" forms, respectively. Also, the SMF part in the proposed hybrid filter processes the observations at a given time sequentially.

For a fair comparison, other filters' configurations follow that of PSMF according to their structures. First, systematic resampling is performed in PF, SIR-ESRF and GMM-EnKF. Second, the same smoothing step is adopted in PF. Third, the map design configurations are the same as in SMF. Finally, the observations are sequentially processed in EnKF, ESRF, SMF, and the ESRF/EnKF parts in hybrid filters.

5.3. Performance evaluation in relation to ensemble size

In the first set of experiments, we numerically investigated the performance of filters under different ensemble sizes over a range of state-space models. The number of particles used was successively $N = [20, 40, 60, 100, 200, 400, 600]$. For the hybrid filters, the parameter θ has influences on their filtering performance. In this experiment, the value of θ was taken from a vector $[0.001, 0.002, 0.004, 0.006, 0.008, 0.01, 0.02, 0.04, 0.06, 0.08, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 0.92, 0.94, 0.96, 0.98, 0.99, 0.992, 0.994, 0.996, 0.998, 0.999]$, and we selected the best performing θ and its corresponding results as the final results. The coordinates in the vector range from 0.001 to 0.999. In the scales near 0 and near 1, a smaller sampling interval is chosen. The reason behind this is that for some of the state-space models, the performance of the hybrid filters is more sensitive to the change of θ within those two scales. The minimum θ in the vector can ensure the effective numbers are less than 1, and maximum larger than $N - 1$, which can ensure the best parameter values are in the tested range. Alternatively, we can use a root finding method to search the best performing hybrid filters and their corresponding values of θ [35]. The root mean-square error (RMSE) was used to analyse the per-
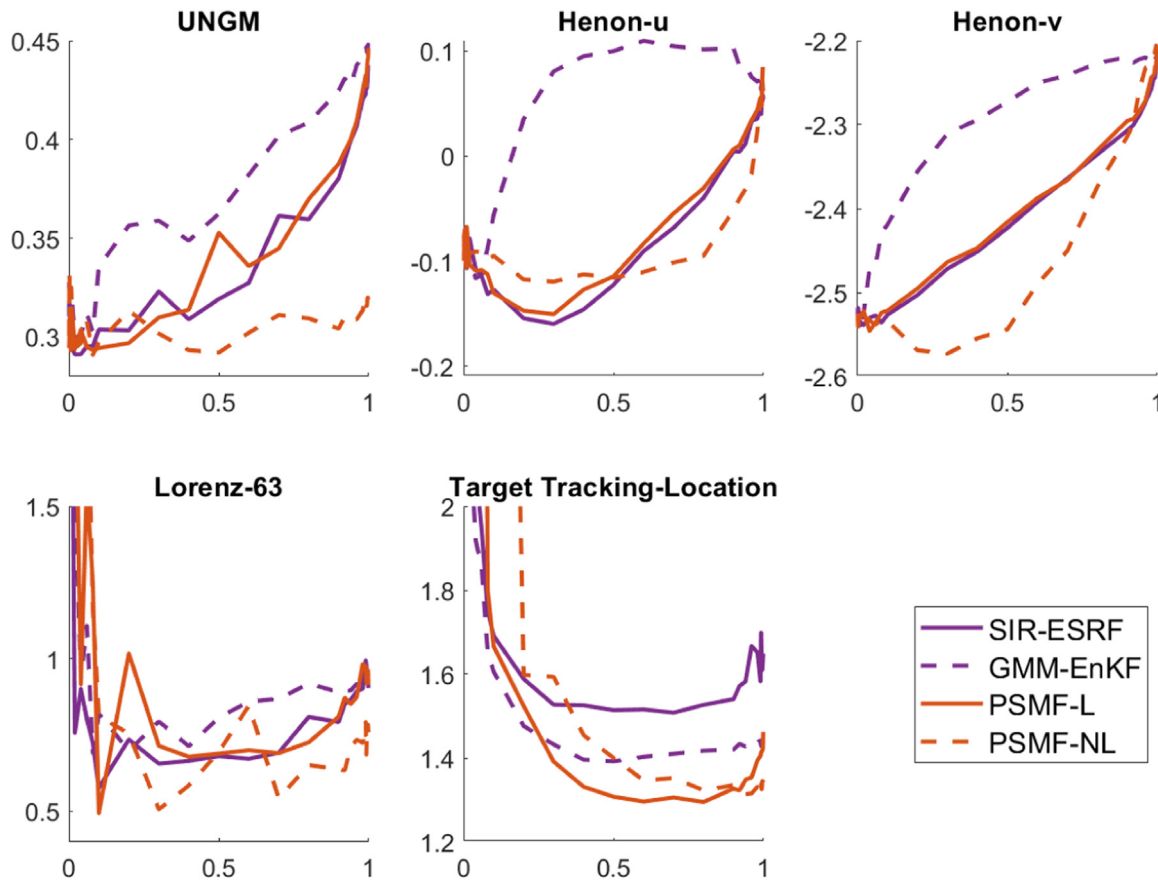
Fig. 7. The comparison among hybrid filters with different values of parameter θ under different models. The ensemble size N is set to 200. Horizontal axis represents the value of θ, and vertical axis represents the estimation log(RMSE).

formance of each filter. For the UNGM and Lorenz-63, estimations on all variables were used to calculate RMSE values. For Henon map model, following [35], the variable u and v were analysed separately whilst the location estimation precision was considered for the target tracking model. In addition, the average continuous ranked probability score (CRPS) is also computed. Different from RMSE, it quantifies the spread of the ensemble. Lower CRPS indicates that the ensemble concentrates around the true values, and is more precise for state estimation.

5.3.1. Evaluating PSMF vs. traditional filters

RMSE comparisons between the PSMF and the traditional filters are plotted in Fig. 2. When compared to the PF for a small number of particles ($N < 100$), the proposed approach provides improved results in all cases. For large ensemble sizes, although PSMF still yields better tracking performance under higher dimensional models, similar estimations are obtained under the one-dimension model. These comparisons indicate that different from the classic PF, the proposed hybrid filter can effectively alleviate particle degeneracy due to employing the parameter θ. Note however that in one-dimensional cases, because of the narrow sampling space, for medium and large number of particles, the classic PF does not suffer from particle degeneracy seriously, and PSMF does not make a significant difference in these cases.

For small N, the PSMF has similar results with the EnKF/ESRF, whereas when increasing the number of particles, the PSMF starts to show considerable gain for models. This shows that with few number of samples, the PSMF utilises a larger θ value to maintain the effective number of particles, so that α tends to be smaller leading to the PSMF performing closely to the SMF. By contrast,

with large N, the PSMF adopts a smaller θ, and α increases. Then, the PF part occupies a larger assimilation proportion, and more nonlinear information can be extracted. On the other hand, due to their limited flexibility, the EnKF/ESRF does not yield better results with increased number of particles.

Specifically speaking for the linear and nonlinear versions of the proposed approach, for small N, the PSMF-L yields better performance compared to the PSMF-NL. In most cases, with an increase in the number of particles, the advantages of the PSMF-L gradually reduces. The reasons behind this are related to the quality of transport maps. From [29], we can recall that a linear map has more robust performance with small N but can have limited flexibility. By contrast, a nonlinear map holds higher non-linearity but might yield unacceptable variances without a sufficient number of particles. As a result, for small N, nonlinear maps cause large variance, and the PSMF-NL performs worse than its linear counterpart, the PSMF-L. For large N, the variance produced by the nonlinear maps can be suppressed and the difference between PSMF-L and PSMF-NL becomes smaller.

The results above do not imply that the PSMF-L can always achieve better results than PSMF-NL. From the results under the variable v of the Henon map, PSMF-NL provides better results. As mentioned before, compared with the nonlinear maps, linear maps have very limited flexibility, which causes poor estimation results. In some cases, a large θ is needed to maintain the effective number of particles, and then the value of α decreases. Consequently, the SMF occupies a larger assimilation part and needs to perform more complex transportations. If the bias caused by the linear transport approximation is larger than that of nonlinear-map variance, the PSMF-NL can perform better than the PSMF-L.
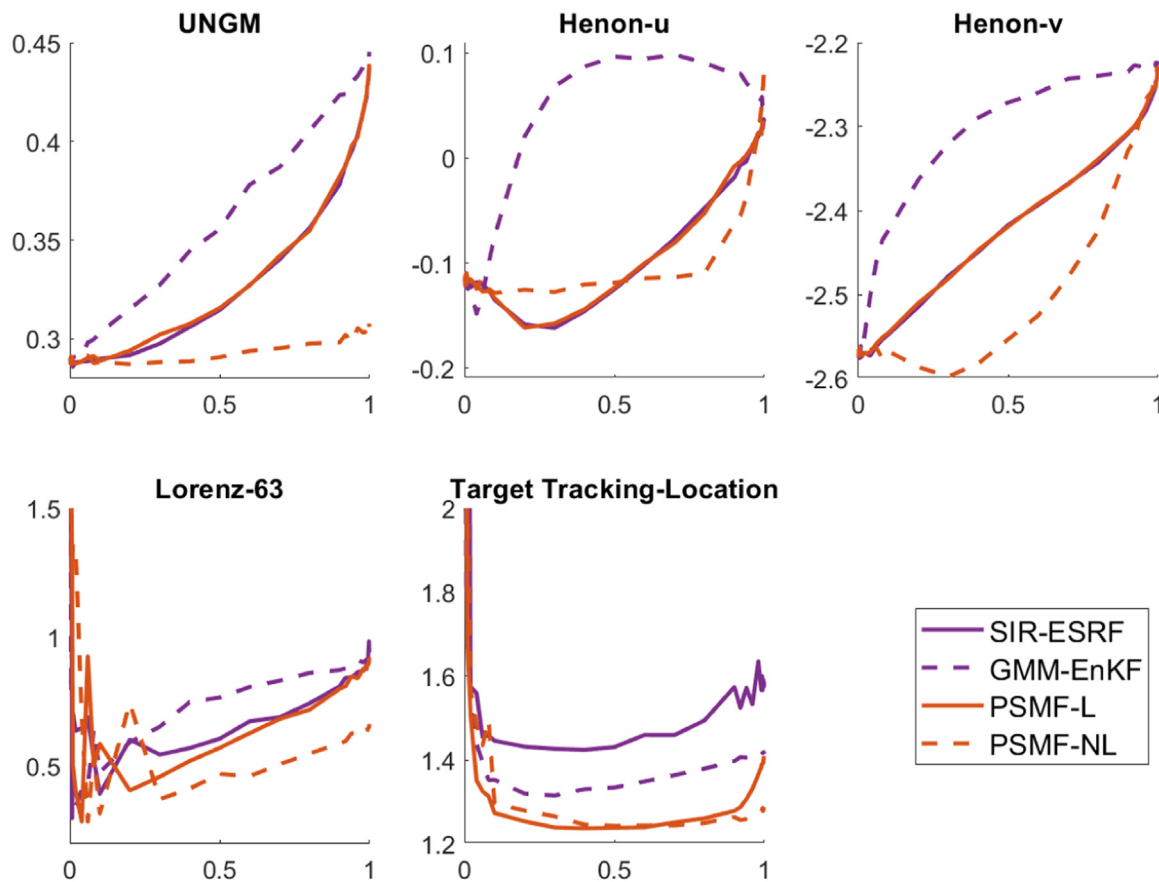
Fig. 8. The comparison among hybrid filters with different values of parameter θ under different models. The ensemble size N is set to 600. Horizontal axis represents the value of θ, and vertical axis represents the estimation log(RMSE).

To complete this section, CRPS comparisons between the PSMF and the traditional filters are also plotted in Fig. 3 and they follow the same characteristics similar to the RMSE results in Fig. 2. This suggests that our new hybrid filter shows advantages over the traditional filters in both mean estimations and ensemble concentration.

5.3.2. PSMF vs. hybrid filters

RMSE comparisons between the PSMF and other hybrid filters are plotted in Fig. 4.

When compared to the SIR-ESRF, the proposed PSMF performs worse when the particle number is small. In most cases, for increasing number of particles, the proposed PSMF's performance achieved is closer to or better than that of SIR-ESRF.

The difference between the performance of the SIR-ESRF and that of PSMF is primarily due to their distinct strategies of alleviating the particle degeneracy. In the PSMF, systematic resampling with a smoothing step is adopted. By contrast, SIR-ESRF relies on the mean-preserving random orthogonal transformation which resamples the ensemble during the ESRF assimilation stage. This resampling technique performs well under medium-level nonlinear models. However, since it is based on the Gaussian approximation, when the posterior distribution gets further from Gaussian, more errors will be introduced. In all the experiments, for small N, the bias caused by the systematic resampling and transport map variances is larger than that of the Gaussian approximation in SIR-ESRF, which thus determines the proposed PSMF to achieve closer or worse results. However, for medium and large N, the bias from the systematic resampling and transport map variance is less sig-

nificant. As a result, it becomes smaller than the error from the resampling in the SIR-ESRF.

For small N, the PSMF and GMM-EnKF achieve considerably lower performance compared to the SIR-ESRF. However, the GMM-EnKF achieves worse results than the proposed PSMF for a medium number of particles. When increasing the number of particles, its performance gradually approaches that of PSMF, although under the nonlinear non-Gaussian target tracking model, for $N = 600$, the PSMF still offers obvious advantages. These results are caused by the structure of the GMM-EnKF. Different from the other hybrid filters, the GMM-EnKF implements the EnKF assimilation first, which is based on the Gaussian assumption for the prior distributions. In all the state-space models in this paper, all the measurement models are linear and Gaussian, which means that the posterior has a closer form to a Gaussian than the prior. Thus, implementing ensemble-based filters first introduces more error. For small N, the EnKF update occupies a large part to keep the *ESS*. As a result, the Gaussian approximation introduces more errors. For large N, the GMM-EnKF just needs a small θ value to ensure the *ESS* is large enough. Consequently, a smaller α will be selected and the GMM occupies a large proportion. Hence, more nonlinear transform information between the prior and posterior can be obtained. However, the 4-dimensional target tracking model has a larger sampling space and needs more effective particles. So, even for $N = 600$, the EnKF occupies a larger assimilation proportion than the other cases. As a result, the GMM-EnKF is significantly outperformed by the proposed PSMF.

As in previous section, we also provide comparisons between the PSMF and the other hybrid filters in terms of CRPS and these are plotted in Fig. 5. The results are consistent with the RMSE
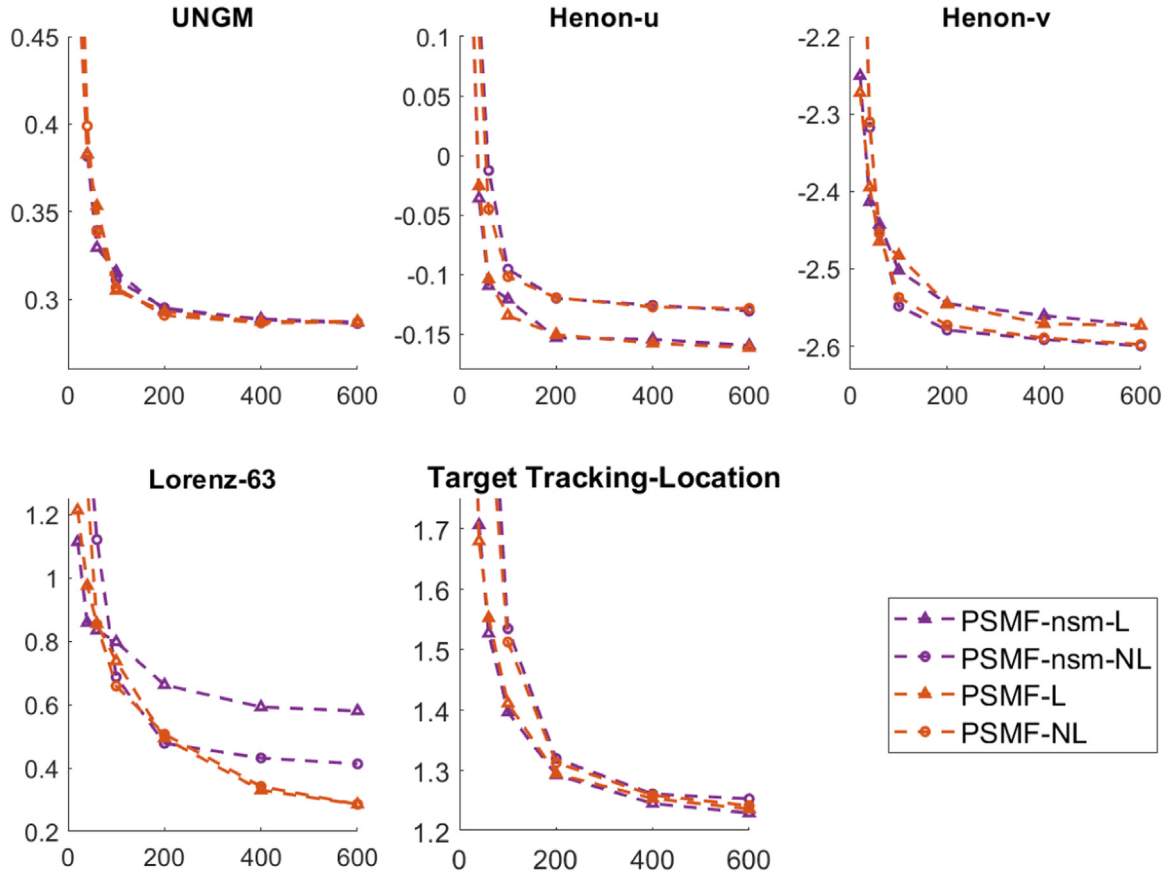
Fig. 9. The comparison of the proposed PSMFs between with and without the smoothing step under the different number of particles. PSMF-nsm-L and PSMF-nsm-NL represent linear and nonlinear PSMFs without the smoothing step, respectively. The horizontal axis represents the number of particles, and the vertical axis represents log(RMSE).

results in Fig. 4. This means that the differences among the hybrid filters in the mean estimation can be extended to the ensemble concentration. Finally, for further objective evaluation, numerical results obtained for both RMSE and CRPS are also shown in Tables A.1 and A.2, respectively, in Appendix A. Due to the limited space, a sparse selection of ensemble sizes, [20,60,200,600], is provided. Results indicate that in most cases, hybrid filters work better than traditional filters. Also, SIR-ESRF tends to perform better than the others with few particles, while PSMF obtains robust performance with large ensemble sizes.

5.4. The effect of θ on filter performance

In the second set of experiments, we investigated the influence of θ on the hybrid filters. We used the same values of θ as in Section 5.3, i.e., between 0.001 and 0.999. The number of particle used were 20, 200, and 600.

In Fig. 6, the influence of parameter θ is evaluated, for an ensemble size of $N = 20$. It can be observed that, in general, the best results are achieved by the SIR-ESRF. As was mentioned in the previous sections, the reason behind this is that the resampling technique of the SIR-ESRF, i.e., the mean-preserving random orthogonal transformation, has a more robust performance for small N. In addition, the large variance of nonlinear maps for small number of particles makes the proposed nonlinear approach be inefficient in this case.

The PSMF-L and GMM-EnKF achieve inconsistent results and lower performance than the SIR-ESRF in most cases. This is due to the known drawbacks of these two filters: when using few particles, the systematic resampling technique in PSMF-L is not as robust as that in SIR-ESRF. By contrast, the GMM-EnKF runs by first performing the EnKF component, which leads to large errors.

In Fig. 7, the influence of the parameter θ is evaluated for $N = 200$. It can be seen that for the first 3 models, for the same reason as above, the GMM-EnKF yields the worst results. It is clear that the proposed PSMF-NL achieves the best performance for large θ values (close to 1). The improved results benefit from the larger flexibility of the nonlinear maps and reduced map variances in the case of a medium number of particles. However, when θ is small, the advantage of PSMF-NL is not obvious anymore, the reason being that in such case, the values of α tend to be large, and the SMF part occupies a small proportion of the assimilation process. Consequently, there is no need for a flexible nonlinear map to complete the assimilation. Also, we should notice that the performance of the PSMF-NL is just slightly worse than that of PSMF-L in this case. The small effective number of particles caused by small θ do not cause large nonlinear map variances. Because of our smoothing strategy, the diversity of particles can be ensured and the variance can be suppressed.

For the first 3 models, the PSMF-L and SIR-ESRF achieve comparable results with lower performance compared to the proposed PSMF-NL. The disadvantages of the SIR-ESRF stem from the Gaussian assumption of the resampling techniques, whilst the disadvantages of the PSMF-L are due to the lower flexibility of linear maps.

For the last model (4D tracking), the results are slightly different. First, the SIR-ESRF performs relatively worse than the proposed PSMF-L since the posterior distributions under this model are far from the standard Gaussian assumption. Second, when θ is close to 1, the advantage of the proposed PSMF-NL can still be seen. Nevertheless, on a large scale, the PSMF-NL performs worse
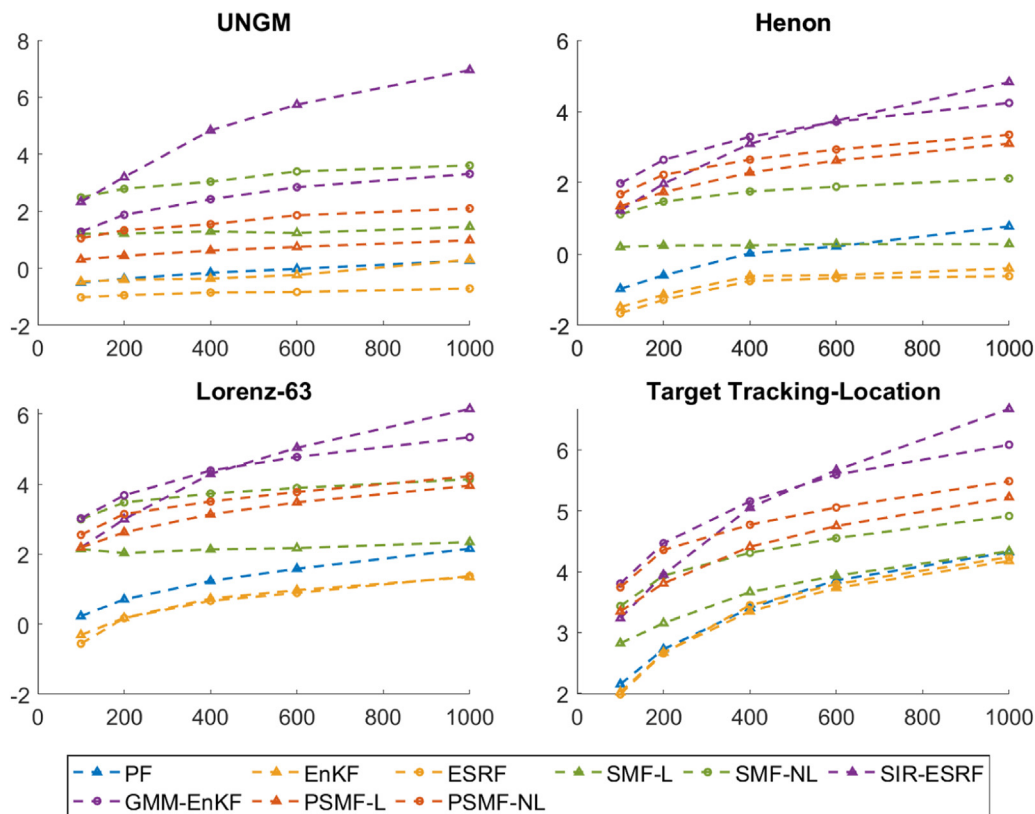
Fig. 10. The comparison of the proposed PSMF to the other filters under the different number of particles. The horizontal axis represents the number of particles, and the vertical axis represents log(time).

than its linear counterpart. This is due to increased variance of the nonlinear map in this case. According to the structure of the triangular maps, it is known that the number of map parameters increases with the dimension of models. As a result, a medium number of particles cannot effectively suppress the variance of nonlinear maps that causes PSMF-NL to perform worse than the linear PSMF-L.

Finally, in Fig. 8, the influence of parameter θ is evaluated, for $N = 600$. The results are similar to those obtained for $N = 200$, with some slight differences. It can be seen that For large N, the proposed PSMF-NL benefits more from the increase of the number of particles. For a larger scale of values of θ, it outperforms the others for the first three models. Even under the 4D target tracking model, the variance of nonlinear maps can be suppressed more efficiently, and its performance becomes closer to that of the PSMF-L.

We can also notice that under the target tracking model, when $0.2 < \theta < 0.8$, the location estimation log(*RMSE*) of the proposed PSMF varies very little. We attribute this to the fact that the transformation between a prior distribution and corresponding posterior distribution is not particularly complex under this model. When θ changes between 0.2 and 0.8, the assimilation proportion of PF changes as well. However, the SMF can always complete the remaining assimilation update and the final estimation results remain stable. When θ is very close to 1, transport maps cannot extract all the remaining nonlinear information and the performance is reduced. By contrast, when θ is very close to zero, the PSMF becomes closer to the classical PF and hence suffers from particle degeneracy.

5.5. The effect of the smoothing step on filter performance

In this section, the influence of the smoothing step on the filtering performance of PSMFs is investigated. The parameter set is the same as in the first experiment, and the experiment results are shown in Fig. 9. Under most models, the smoothing step does not influence the filtering performance. However, under Lorenz-63, with a large number of particles, PSMF performs better than PSMF-nsm. The reason is that with large ensemble sizes, the optimal value of θ is relatively small under Lorenz-63 as shown in Fig. 8. Thus, *ESS* is small, which causes large map variances. A smoothing step can increase the diversity of samples, and then the map variance is decreased. Hence, in this case, PSMF works better than PSMF-nsm.

5.6. Computational complexity analysis

For an exhaustive evaluation of our proposed PSMF method, the filter efficiency with the change of particle numbers is also evaluated. To this end, we fix the parameter θ to 0.5, and employ ensemble sizes of $[100, 200, 400, 600, 1000]$. Small particle numbers are not considered because some filters diverge with few samples and then the precise computational cannot be obtained. The rest of the parameters are the same as in the first set of experiments presented above. The computational time is shown in Fig. 10.

Compared with traditional methods, PSMF is less efficient than PF, EnKF and ESRF. However, the comparison between PSMF and SMF does not lead to consistent conclusions, being dependent on the state-space models investigated. Specifically, under the one-dimensional UNGM model, the PF part of PSMF can keep the *ESS* high, and the SMF part is omitted in many iterations. Consequently, PSMF is more efficient than SMF. By contrast, under higher-dimensional models, the SMF part of PSMF can be omitted in fewer iterations. Thus, the inclusion of both PF and SMF steps in PSMF slows the filtering process.

Also, GMM-EnKF is more time-consuming than PSMF for all the models. This is due to its *ESS* calculation for searching α. Specifi-

cally, because GMM-EnKF implements EnKF first, the whole hybrid filter should be implemented to calculate *ESS*. However, for PSMF, only the PF part is implemented to obtain *ESS*, as the SMF step does not change *ESS*. Therefore, GMM-EnKF spends more time on the search of α by the root-finding method.

Finally, PSMF has a similar computational load to SIR-ESRF when the ensemble size is small. However, the computational time of SIR-ESRF increases faster than PSMF along the *x*-axis and becomes larger than that of PSMF when particle numbers are large. This is caused by the mean-preserving random orthogonal transformation in SIR-ESRF, which has nonlinear computational complexity with respect to the particle number as explained in Grooms and Robinson [35]. By contrast, PSMF has linear computational complexity.

6. Conclusion

In this paper, we proposed a novel hybrid filtering approach, the PSMF, which enhances the standard particle filter by using ideas from stochastic map filters. To break the particle degeneracy issue, which is peculiar to PF, systematic resampling followed by a smoothing step was adopted. In order to analyse the impact of the nonlinearity of transport maps, we presented PSMF-L and PSMF-NL by adopting linear and nonlinear transport maps, respectively.

Two sets of experiments involving four widely employed state-space models were implemented to validate the proposed hybrid filters. In the first experimental setup, we investigated the performance of the PSMF with optimal parameters of θ for different ensemble sizes. Experiment results showed that for medium and large number of particles the linear version of the proposed hybrid filter, PSMF-L, yields better results than the benchmark approaches. Nevertheless, in some cases, the error inherently introduced by the nonlinear map variance was smaller than that due to the linear approximation of the linear map. Consequently, the nonlinear hybrid filter provided better results than the PSMF-L. We should also note that, with a small number of particles, the advantages that

the PSMF offers were not obvious when compared to the other filters, since the PSMF suffers errors caused by the systematic resampling and transport map variance.

For the second series of simulations, the relationship between the performance of the PSMF and the parameter θ was investigated. It was demonstrated that in most cases the PSMF-NL is more tolerant to changes of the parameter θ for medium and large numbers of particles. This owes to the higher flexibility of nonlinear maps. In addition, for small numbers of particles, due to the large nonlinear map variance, the PSMF-NL achieves lower performance with respect to benchmark approaches.

Our future work will consider extending the proposed approaches to higher dimensional problems. Moreover, we will also investigate the deterministic transport map filter (DMF) [29] within the context of the proposed hybrid filters.

Data availability

All the used codes for the above experiments are shared at: https://github.com/PengchengH/A-hybrid-particle-stochastic-map-filter.git.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT authorship contribution statement

Pengcheng Hao: Methodology, Software, Writing – original draft. **Oktay Karakuş:** Supervision, Writing – review & editing. **Alin Achim:** Supervision, Writing – review & editing.

Appendix A. Numerical results

Table A1

RMSE results of the filters under different models. R and N represent RMSE and the ensemble size, respectively. M1:M5 represent UNGM, Henon-u, Henon-v, Lorenz-63, and Target tracking, respectively. F1:F9 represent PF, EnKF, ESRF, SMF-L, SMF-NL, SIR-ESRF, GMM-EnKF, PSMF-L, PSMF-NL. The first and second minimums under different particle numbers and models are highlighted in red and blue bold font, respectively.

R	N	F1	F2	F3	F4	F5	F6	F7	F8	F9
M1	20	2.375	1.81	1.681	1.733	1.876	1.576	1.647	1.651	1.778
	60	1.536	1.598	1.641	1.59	1.529	1.398	1.422	1.424	1.403
	200	1.351	1.572	1.628	1.574	1.366	1.338	1.346	1.341	1.338
	600	1.335	1.562	1.624	1.564	1.346	1.331	1.334	1.333	1.332
M2	20	1.821	1.446	1.331	1.38	2.008	1.227	1.323	1.239	1.662
	60	1.103	1.14	1.127	1.135	1.239	0.888	0.991	0.901	0.956
	200	0.915	1.057	1.055	1.063	1.094	0.852	0.89	0.86	0.887
	600	0.891	1.043	1.038	1.042	1.121	0.85	0.862	0.851	0.879
M3	20	0.155	0.128	0.119	0.129	0.25	0.102	0.114	0.103	0.155
	60	0.097	0.113	0.112	0.113	0.132	0.086	0.091	0.085	0.086
	200	0.08	0.108	0.109	0.11	0.11	0.079	0.079	0.078	0.076
	600	0.077	0.108	0.108	0.108	0.108	0.076	0.076	0.076	0.074
M4	20	20.361	3.323	4.451	3.667	NaN	2.681	2.977	3.361	9.33
	60	20.229	2.815	4.214	2.872	3.198	2.228	2.46	2.35	2.346
	200	17.282	2.506	4.383	2.673	2.11	1.782	1.997	1.639	1.66
	600	13.533	2.511	3.671	2.477	1.88	1.348	1.409	1.33	1.33
M5	20	996.963	7.788	6.745	8.551	59.663	8.549	8.136	9.214	912.3
	60	690.83	4.856	4.413	4.979	7.857	5.14	4.853	4.723	6.756
	200	458.637	4.207	4.127	4.196	3.821	4.519	4.024	3.652	3.719
	600	262.829	4.116	4.096	4.122	3.655	4.153	3.726	3.445	3.465

Table A2

CRPS results of the filters under different models. C and N represent CRPS and the ensemble size, respectively. M1:M5 represent UNGM, Henon-u, Henon-v, Lorenz-63, and Target tracking, respectively. F1:F9 represent PF, EnKF, ESRF, SMF-L, SMF-NL, SIR-ESRF, GMM-EnKF, PSMF-L, PSMF-NL. The first and second minimums under different particle numbers and models are highlighted in red and blue bold font, respectively

C	N	F1	F2	F3	F4	F5	F6	F7	F8	F9
M1	20	1.031	0.913	0.892	0.898	0.908	0.82	0.843	0.826	0.855
	60	0.767	0.837	0.876	0.834	0.774	0.734	0.747	0.745	0.739
	200	0.714	0.823	0.87	0.822	0.727	0.713	0.716	0.714	0.713
	600	0.709	0.816	0.868	0.816	0.716	0.707	0.708	0.708	0.708
M2	20	1.427	0.878	0.847	0.847	1.129	0.736	0.812	0.743	1.002
	60	0.743	0.669	0.708	0.663	0.717	0.505	0.594	0.513	0.554
	200	0.544	0.609	0.66	0.61	0.629	0.481	0.516	0.486	0.503
	600	0.509	0.597	0.647	0.596	0.647	0.479	0.491	0.478	0.496
M3	20	0.118	0.075	0.07	0.076	0.136	0.059	0.066	0.06	0.087
	60	0.065	0.065	0.065	0.066	0.076	0.049	0.054	0.048	0.048
	200	0.046	0.062	0.062	0.063	0.063	0.045	0.045	0.045	0.042
	600	0.043	0.062	0.062	0.061	0.061	0.042	0.042	0.042	0.041
M4	20	9.28	0.954	1.235	0.987	NaN	0.781	0.871	0.939	2.102
	60	9.355	0.82	1.198	0.834	0.794	0.603	0.688	0.644	0.654
	200	6.99	0.766	1.242	0.787	0.582	0.442	0.484	0.423	0.447
	600	4.466	0.762	1.116	0.757	0.542	0.37	0.36	0.357	0.358
M5	20	378.726	2.738	2.474	2.773	14.909	2.83	2.73	3.043	338.948
	60	224.118	1.844	1.871	1.857	2.429	1.895	1.843	1.8	2.315
	200	97.731	1.673	1.808	1.668	1.518	1.686	1.564	1.438	1.472
	600	36.274	1.64	1.799	1.643	1.449	1.589	1.459	1.362	1.374

References

[1] G.Y. Kulikov, M. Kulikova, Estimation of maneuvering target in the presence of non-Gaussian noise: a coordinated turn case study, Signal Process. (145) (2018) 241–257, doi:10.1016/j.sigpro.2017.12.012.

[2] Y. Huang, Y. Zhang, P. Shi, Z. Wu, J. Qian, J.A. Chambers, Robust Kalman filters based on Gaussian scale mixture distributions with application to target tracking, IEEE Trans. Syst., Man, Cybern. 49 (10) (2019) 2082–2096, doi:10.1109/TSMC.2017.2778269.

[3] P. Dong, Z. Jing, H. Leung, K. Shen, J. Wang, Student-t mixture labeled multi-bernoulli filter for multi-target tracking with heavy-tailed noise, Signal Process. 152 (2018) 331–339, doi:10.1016/j.sigpro.2018.06.014.

[4] R.E. Kalman, A new approach to linear filtering and prediction problems, Trans. ASME–J. Basic Eng. 82 (7) (1960) 35–45, doi:10.1115/1.3662552.

[5] S. Julier, J.K. Uhlmann, A general method for approximating nonlinear transformations of probability distributions, Citeseer, 1996.

[6] S.J. Julier, J.K. Uhlmann, A new extension of the Kalman filter to nonlinear systems, Signal Process., Sens. Fusion, Target Recognit. VI 3068 (1997) 182–193, doi:10.1117/12.280797.

[7] G. Wang, Y. Zhang, X. Wang, Iterated maximum correntropy unscented Kalman filters for non-Gaussian systems, Signal Process. 163 (2019) 87–94, doi:10.1016/j.sigpro.2019.05.015.

[8] A. Doucet, S. Godsill, C. Andrieu, On sequential Monte Carlo sampling methods for Bayesian filtering, Stat. Comput. 10 (3) (2000) 197–208, doi:10.1023/A:1008935410038.

[9] M.S. Arulampalam, S. Maskell, N. Gordon, T. Clapp, A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking, IEEE Trans. Signal Process. 50 (2) (2002) 174–188, doi:10.1109/78.978374.

[10] Z. Chen, Bayesian filtering: from Kalman filters to particle filters, and beyond, Statistics 182 (1) (2003) 1–69, doi:10.1080/02331880309257.

[11] T. Bengtsson, P. Bickel, B. Li, Curse-of-dimensionality revisited: collapse of the particle filter in very large scale systems, Inst. Math. Stat. Collect. 2 (1) (2008) 316–334, doi:10.1214/193940307000000518.

[12] X. Wang, T. Li, S. Sun, J.M. Corchado, A survey of recent advances in particle filters and remaining challenges for multitarget tracking, Sensors 17 (12) (2017) 2707, doi:10.3390/s17122707.

[13] N. Branchini, V. Elvira, Optimized auxiliary particle filters: adapting mixture proposals via convex optimization (2021). 10.48550/arXiv.2011.09317

[14] J. Luo, Z. Wang, Y. Chen, M. Wu, Y. Yang, An improved unscented particle filter approach for multi-sensor fusion target tracking, Sensors 20 (23) (2020) 6842, doi:10.48550/arXiv.2011.09317.

[15] V. Elvira, L. Martino, D. Luengo, M.F. Bugallo, Heretical multiple importance sampling, IEEE Signal Process. Lett. 23 (10) (2016) 1474–1478, doi:10.1109/LSP.2016.2600678.

[16] T. Li, M. Bolic, P.M. Djuric, Resampling methods for particle filtering: classification, implementation, and strategies, IEEE Signal Process. Mag. 32 (3) (2015) 70–86, doi:10.1109/MSP.2014.2330626.

[17] J.D. Hol, T.B. Schon, F. Gustafsson, On resampling algorithms for particle filters, 2006 IEEE Nonlinear Statistical Signal Processing Workshop, 2006, doi:10.1109/NSSPW.2006.4378824.

[18] J. Read, K. Achutegui, J. Míguez, A distributed particle filter for nonlinear tracking in wireless sensor networks, Signal Process. 98 (2014) 121–134, doi:10.1016/j.sigpro.2013.11.020.

[19] T. Li, T.P. Sattar, S. Sun, Deterministic resampling: unbiased sampling to avoid sample impoverishment in particle filters, Signal Process. 92 (2012) 1637–1645, doi:10.1016/j.sigpro.2011.12.019.

[20] G. Evensen, Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistic, J. Geophys. Res. 99 (C5) (1994) 10,143–10,162, doi:10.1029/94JC00572.

[21] G. Burgers, P.J. van Leeuwen, G. Evensen, Analysis scheme in the ensemble Kalman filter, Mon. Weather Rev. 126 (1998) 1719–1724, doi:10.1175/1520-0493(1998)126<1719:ASITEK>2.0.CO;2.

[22] N. Papadakis, E. Mémin, A. Cuzol, N. Gengembre, Data assimilation with the weighted ensemble Kalman filter, Tellus A 62 (5) (2010) 673–697, doi:10.1111/j.1600-0870.2010.00461.x.

[23] M.K. Tippett, J.L. Anderson, C.H. Bishop, T.M. Hamill, J.S. Whitaker, Ensemble square root filters, Mon. Weather Rev. 131 (7) (2003) 1485–1490, doi:10.1175/1520-0493(2003)131<1485:ESRF>2.0.CO;2.

[24] P.J. van Leeuwen, H.R. Künsch, L. Nerger, R. Potthast, S. Reich, Particle filters for high-dimensional geoscience applications: a review, Q. J. R. Meteorol. Soc. 145 (723) (2019) 2335–2365, doi:10.1002/qj.3551.

[25] M. Pulidoab, P.J. van Leeuwen, Sequential Monte Carlo with kernel embedded mappings: the mapping particle filter, J. Comput. Phys. 396 (2019) 400–415, doi:10.1016/j.jcp.2019.06.060.

[26] C.-C. Hu, P.J. van Leeuwen, A particle flow filter for high-dimensional system applications, Q. J. R. Meteorol. Soc. 147 (737) (2021) 2352–2374, doi:10.1002/qj.4028.

[27] S. Kolouri, S.R. Park, M. Thorpe, D. Slepcev, G.K. Rohde, Optimal mass transport: signal processing and machine-learning applications, IEEE Signal Process. Mag. 34 (4) (2017) 43–59, doi:10.1109/MSP.2017.2695801.

[28] S. Reich, A nonparametric ensemble transform method for Bayesian inference, SIAM J. Sci. Comput. 35 (4) (2013) A2013–A2024, doi:10.1137/130907367.

[29] A. Spantini, R. Baptista, Y. Marzouk, Coupling techniques for nonlinear ensemble filtering(2019). 10.48550/arXiv.1907.00389

[30] N. Bonnotte, From Knothe's rearrangement to Brenier's optimal transport map, SIAM J. Math. Anal. 45 (1) (2013) 64–87, doi:10.1137/120874850.

[31] Y. Marzouk, T. Moselhy, M. Parno, A. Spantini, Sampling via measure transport: an introduction, Handbook of Uncertainty Quantification, 2016, doi:10.1007/978-3-319-11259-6-23-1.

[32] A.S. Stordal, H.A. Karlsen, G. Nævdal, H.J. Skaug, B. Vallès, Bridging the ensemble Kalman filter and particle filters: the adaptive Gaussian mixture filter, Comput. Geosci. 15 (2011) 293–305, doi:10.1007/s10596-010-9207-1.

[33] M. Frei, H.R. Kunsch, Bridging the ensemble Kalman and particle filters, Biometrika 100 (4) (2013) 781–800, doi:10.1093/biomet/ast020.

[34] N. Chustagulprom, S. Reich, M. Reinhardt, A hybrid ensemble transform particle filter for nonlinear and spatially extended dynamical systems, SIAM/ASA J. Uncertain. Quantif. 4 (2016) 592–608, doi:10.1137/15M1040967.

[35] I. Grooms, G. Robinson, A hybrid particle-ensemble Kalman filter for problems with medium nonlinearity, PLoS One (2021), doi:10.1371/journal.pone.0248266.

[36] C. Villani, Optimal Transport, Old and New, Springer, Berlin Heidelberg New York, 2008.

[37] T.A.E. Moselhy, Y.M. Marzouk, Bayesian inference with optimal maps, J. Comput. Phys. 231 (23) (2012) 7815–7850, doi:10.1016/j.jcp.2012.07.022.

[38] M.D. Parno, Y.M. Marzouk, Transport map accelerated Markov chain Monte Carlo, SIAM/ASA J. Uncertain. Quantif. 6 (2) (2018) 645–682, doi:10.1137/17M1134640.

[39] C. Cotter, S. Cotter, P. Russell, Ensemble transport adaptive importance sampling, SIAM/ASA J. Uncertain. Quantif. 7 (2) (2019) 444–471, doi:10.1137/17M1114867.

[40] J. Poterjoy, A localized particle filter for high-dimensional nonlinear systems, Mon. Weather Rev. 144 (1) (2016) 59–76, doi:10.1175/MWR-D-15-0163.1.

[41] S. Robert, H.R. Künsch, Localizing the ensemble Kalman particle filter, Tellus A 69 (1) (2017), doi:10.1080/16000870.2017.1282016.

[42] L. Mihaylova, P. Brasnett, A. Achim, D. Bull, N. Canagarajah, Particle filtering with alpha-stable distributions, IEEE/SP 13th Workshop on Statistical Signal Processing, 2015, doi:10.1109/SSP.2005.1628625.

[43] D. Henke, C. Magnard, M. Frioud, D. Small, E. Meier, M. Schaepman, Moving-target tracking in single-channel wide-beam SAR, IEEE Trans. Geosci. Remote Sens. 50 (11) (2012) 4735–4747, doi:10.1109/TGRS.2012.2191561.

[44] D. Henke, E.M. Dominguez, D. Small, M.E. Schaepman, E. Meier, Moving target tracking in single- and multichannel SAR, IEEE Trans. Geosci. Remote Sens. 53 (6) (2015) 3146–3159, doi:10.1109/TGRS.2014.2369060.